LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-582572

WaterMap and Its Implementation in Drug Discovery

Y. Yang, S. E. Wong, F. C. Lightstone

September 18, 2012

Expert Opinion on Drug Discovery

Disclaimer

WaterMap and Its Implementation in Drug Discovery

Yue Yang, Sergio E. Wong, Felice C. Lightstone

Biosciences and Biotechnology Division, Lawrence Livermore National Laboratory, Livermore, CA, USA

1. Introduction

Solvents play quite an important role in most chemical and biological processes[1-3]. It is widely accepted that the presence of water or other solvents in many chemical reactions can result in much lower energy barrier. In enzymatic catalysis, water mediate reaction pathways have been observed in various studies[4,5]. In addition, different conformation flexibility and hydrogen bond patterns have been discovered for cyclic peptides in the presence of membrane and water[6], further illustrating the impact of solvent in biological activities such like membrane penetration. Moreover, as will be discussed later in this review, water also plays a critical role in host-guest chemistry and thus is essential to drug design[7,8]. As such, it is not surprising that accounting for solvents is critical in drug discovery since drugs must modulate biological systems.

Solvent interactions have an impact in all stages of drug discovery. During early discovery stages water solubility is necessary for testing weak inhibitors and subsequent optimization i.e. poorly soluble compounds easily precipitate making testing difficult, if not impossible. In terms of lead optimization, ligand and target interactions with water are both important. For the ligand, Lipinski's rule of 5[9] tells us a logP (octanol-water partition coefficient) below 5 is desirable for an orally active compound. For the target, displacement of weakly-bound structural waters by a ligand can sometimes improve binding affinity[8]; alternatively strongly-bound structural waters can serve as hydrogen bond donors or acceptors for an inhibitor. It is noteworthy that while a compound's chemical structure is not mutable in the later stages of drug discovery, its physico-chemical properties have a profound effect on its success. The ability of the compound to penetrate the gut-wall, enter the blood stream, permeate through cell membranes and, sometimes, the nuclear membrane hinges on a balance between interactions with polar and non-polar environments. For all these reasons, water is the most important solvent in drug discovery.

A proper description of solvation effects has always been a challenge in molecular modeling and simulations[3,10]. Many approaches have been made to capture solvents effects in biological processes, by either applying continuum (implicit) solvent models[11-16], explicit solvent models[17-20], or treating solvent molecules quantum mechanically[21-23].

Explicit solvent models, as the name indicates, treat solvents explicitly based on molecular mechanics approximations, which benefits the reproduction of solvent molecules' physical function in the focused processes. Different water models have been developed[18,24], such as the widely implemented TIP3P, TIP4P and finer models. However, more computational cost would be expected associated with more advanced models. Implicit solvent models treat solutions as continuum dielectrics based on the Poisson-Boltzmann (PB) equation. Kuhn and Kollman conducted a famous binding free energies analysis using PB solvation model[14], making it a popular approach thereafter. The potential high computational cost associated with solving PB equation stimulates the development of Generalized Born (GB) approach, which is still based on PB equation but using more approximations, at fairly high accuracy. In fact, the PB and GB based MM-PB/SA and MM-GB/SA has become very popular approaches nowadays in drug discovery. In spite of numerous solvent models developed, it is still difficult to capture solvation effects accurately[25] even when using extensive free energy calculations in the context of statistical mechanics. Efforts to develop solvent models and strengthen understanding of solvent effects in chemical or enzymatic processes have never stopped[23,26,27]. Among all of the solvents, water represents a simple but challenging and somewhat critical model[28,29]. On one hand, water is essential to the function of most biological complexes. They are capable of not only stabilizing structures in protein folding[2,30-32] or/and ligand-protein binding, but also often participating in catalysis[1,5,33,34]. On the other hand, despite many different implicit or explicit models, or even quantum (some times *ab initio*) approaches, it remains challenging to reproduce its physical properties.

Computer-aided drug design (CADD) aims to design ligands that bind specific target proteins with high affinities. The proper treatment of water in drug discovery is essential but complicated because: 1) desolvation of the protein binding pocket and ligand both contribute to the binding affinity and 2) the solvent dielectric dampens electrostatic interactions in a non-uniform manner that depends on the specific geometry of protein/ligand complex. There are many reviews focused on this topic[27,29,35-38]. To further complicate matters, structural waters often bridge or accompany ligand-protein interactions. An average of 4.6 ligand-bound water molecules were identified in each

ligand-receptor complex according to a recent survey[39]. Unfortunately, not all crystallographic waters behave the same way. The displacement of some crystal waters by a ligand often is entropically favorable, but that is not always the case[8]. The complex energetics of structural waters displacement complicate docking calculations that aim to predict binding poses and estimate binding affinities. To understand the role of a specific water molecules present in or near the binding site is important but hard to achieve, since such attempt usually associated with costly statistical mechanics based free energy approaches. Major efforts have been made to improve water/solvent effect in CADD in two directions: (1) to handle waters in docking and virtual screening calculations, and (2) to predict water displacement in the lead optimization stage[40]. Many studies focused in the virtual screening aspects and improvements were reported[41-49], while the latter is a trickier task since the water network might be totally different in apo structures and in ligands bound complexes. In the thermodynamic decomposition of the ligand-protein binding processes, the desolvation of the ligand-complementary protein binding pocket usually accounts for a non-trivial part, if not the dominant part, of the total binding affinity. Therefore, accurately describing water displacement is critical to predicting binding affinities. Unfortunately, traditional approaches are not helpful in this regard – most X-ray structures detect only a subset of waters without any energetic or thermodynamic information, while those popular computational methods applied in CADD, such as MM-PB/SA and/or MM-GB/SA, tend to ignore this entropy contribution. However, computational chemists still prefer MM-PB/SA and/or MM-GB/SA because of the compromise between speed and accuracy.

Higher accuracy can be achieved via computationally more expensive statistical mechanics based free energy approaches. These methods, including thermodynamic integration (TI) or free energy perturbation theory (FEP), can help detect the loci and energetics or water molecules in the binding pocket. Pan et al. reported conducting grand canonical ensemble free energy simulations to improve the potency of lead compounds by targeting weakly bound water at the ligand-protein interface[50]. Yu and Rick implemented thermodynamic integration molecular dynamics[51] (MD) simulations to estimate the free energy, entropy and enthalpy of binding a water molecule in DNA gyrase[52] and other proteins[53]. Their studies revealed the unfavorable entropy (-1.1 to -3.1

kcal/mol for $T\Delta S$[52]) associated with the addition of water to the binding sites, and a good correlation between the decreasing free energy and the change of hydrogen bond numbers associated with water addition. Baron, Setny and McCammon investigated water thermodynamics associated with several cavity-ligand recognitions based on a potential of mean force (PMF) approach, and discovered the main driving force of molecular recognition or rejection is enthalpy or entropy contributions from water instead of ligand or receptor[54]. Later, Baron and Molinero extended this approach to Coarse-grained simulation and made comparison to atomistic simulations[55]. Michel, Tirado-Rives and Jorgensen performed a Monte Carlo (MC) based free energy perturbation (FEP) simulation coupled with JAWS (Just Add Water Molecules, an approach designed to predict the water placement in binding sites for a given structure[56]) to inspect the impact of displacing ordered water in the active sites of three enzymes[25,56]. Their results strongly suggested that careful examination and analysis is required for accurate results, as direct modification of ligand in free energy calculation might result in trapping ordered water thus misleading the lead optimization. A subsequent research by Jorgensen and coworkers illustrated that the traditional solvation approach of immersing the protein/ligand system in a water box affects the accuracy and the combination of the JAWS algorithm and MC/FEP can dramatically improve the results, since the initial placement of water molecules play a critical role in this protocol[57]. In addition, Essex and coworkers have carried out MC/FEP protocol in the context of replica exchange simulation to examine the difficulty level of displacing a conserved or a displaceable water molecule in the active site of a few ligand-protein complexes[58], and the free energy changes (ΔG) of -2.2, -2.4 and +3.0 kcal/mol were reported. Conceivably, significantly lower binding free energies were reported for those displaceable waters than those conserved. Moreover, Michel and Essex also reviewed several popular free energy approaches in this regard and to the extent of predicting absolute protein-ligand binding affinities[59]. Interestingly, this review also noted that a computed free energy gain of 1.0 kcal/mol or more is more likely to be detected in experimental measurements.

Unfortunately, the extensive time-demanding nature of these methods prevents them from being widely applied in CADD, which usually requires the scaning of tens to thousands of lead compounds. To fill the gap between accuracy and speed, WaterMap

has become a popular subject in recent years[7,8,60]. Successful implementations of WaterMap to improve virtual screening and/or lead optimization have been reported. Though it is neither a robust tool that predicts standard (absolute) binding free energies nor a very cheap approach due to the use of MD simulations, it works amazingly well on congeneric ligands with hydrophobic effect dominant binding processes. In this review, we will focus on WaterMap and introduce the basics and some recently notable successes in guest-host drug discovery with this protocol applied. Readers, after reading this review, will become familiar with the basic ideas, implementation, limitations and potential improvements of this protocol, as well as the strategy to face water displacement in CADD.

2. Basic concept of WaterMap

Hydrophobic interactions are believed to deliver the principal thermodynamic driving force to molecular recognition and the binding process. The idea underlying this assumption is that a system gains in binding free energy via releasing molecules from a hydrophobic thus suboptimal pocket into bulk solvent.

WaterMap[7,8,61] is a protocol that performs a post MD trajectory-analysis based on inhomogeneous solvation theory proposed by Lazaridis in 1998[62,63] to calculate the free energy cost of moving a water molecule from a protein hydration site into bulk solvent. The inhomogeneous solvation theory features an entropy expansion term as the function of orientational and spatial particle correlations. In this model, bulk solvent corresponds to zero entropy while the excess entropy can be estimated for structured water localized in or near the binding site. In the application of the WaterMap protocol, the first order and partial of the second order of the expansion will be calculated (Equation 1).

$$S_e = -\frac{k_b \rho_\omega}{\Omega} \int g_{sw}(\mathbf{r}, \omega) \ln g_{sw}(\mathbf{r}, \omega) d\mathbf{r} d\omega$$
$$-\frac{k_b \rho_\omega^2}{2\Omega^2} \int g_{sw,w}(\mathbf{r}^2 \omega^2) \ln \delta g_{sw,w}(\mathbf{r}^2, \omega^2) d\mathbf{r}^2 d\omega^2 - \cdots,$$

Equation 1. Excess entropy estimated by inhomogeneous solvation theory.

where **r** and ω are Cartesian coordinates and Euler angle orientation of water, $g_{sw}(\mathbf{r},\omega)$ describes the single-body distribution of water at **r** and ω, $g_{sww}(\mathbf{r},\omega)$ gives the two-body distribution and ρ_w corresponds to the density of the bulk.

Although fairly accurate results have been achieved by using traditional approaches that rely on pairwise atom-atom or buried surface area terms parameterized against different bio-environments to describe hydrophobic effect, such a strategy fails to explain the super affinity in the well-known streptavidin-biotin complex. Young et al conducted the first WaterMap application[7] in order to address this problem. Short (10ns) MD simulations were carried out using the OPLS-2005 all-atom force filed[64,65] and TIP4P[18] water model on several apo protein receptors. The non-solvent heavy atoms were harmonically restrained, which mimics the so-called rigid receptor approximation widely implemented in molecular docking. A subsequent cluster analysis was performed to identify the principal hydration sites, through partitioning the binding-cavity solvent density distribution. The next step is to estimate the entropy penalty of structure water at each hydration sites using inhomogeneous solvation theory[62,63]. A stable five-membered water ring in the binding site of streptavidin was highlighted and the corresponding entropic contribution (~ -7 kcal/mol) to the free energy of water displacement was estimated to contribute up to five orders of magnitude in the binding affinity constant. On the contrary, solvation of COX-2 binding cavity was found energetically unfavorable and it was believed the narrow hydrophobic enclosure discourages the formation of complementary hydrogen bonds. Certain mutations capable to remove hydrophobic enclosure could result in more entropically favorable solvation in the binding site. WaterMap powered explicit solvent simulations successfully explained these binding behaviors, nonetheless PB-based methods underestimates the binding affinities because of their incapability on molecular-length scale solvation physics.

Abel et al. extended the effort to factor Xa (fXa) and built a model to predict the effect of displacing water from the active site with atomic detail[8]. Using data from short MD simulations, a total of 31 pairs of fXa ligands were examined and a high correlation with experimental relative binding free energies (R^2=0.81, or 0.80 using leave-one-out validation) was obtained. Comparing to a correlation of R^2=0.29 from a traditional MM-GB/SA approach, the advantage and efficiency of the WaterMap protocol is obvious. In

addition, the hydration sites detected by WaterMap were directly compared to the crystallographic water locations[66] as well as data from previous MD study[67], and showed high consistency (9 of 11 crystal water molecules are within 1.5Å from predicted hydration sites, while 2 others are also within 2.5Å range), except for those detected by computational procedure but not present in crystal structure (1HCG, resolved at 2.2Å). This discrepancy is possibly attributed to the sensitivity of identifying water molecules to the resolution of X-ray crystallography[68] and many other physical factors, such as temperature, pH and so on.[69,70] More importantly, a descriptors (Equation 2) was developed to quantitatively predict the impact of displacing water in the binding site. Such a scoring function successfully captured the difference in free energy contribution of substituting solvent, energetically favorable or unfavorable, for the congeneric sets of ligand. Comparing to traditional approaches in which a specific water molecule is considered either energetically favorable or unfavorable to be replaced, WaterMap protocol investigates whether displacing a water molecule is thermodynamically favorable or not, therefore producing more reliable prediction in binding affinity differences. Computational cost of this work is also modest - meaningfully only one MD simulation was performed on a single structure, which was subsequently used for the superposition of congeneric ligands. However, fXa is a receptor well known for the super rigidity, therefore, the real computational expense for a comparable size of study on another protein might be much higher.

$$\Delta G_{bind} = \sum_{lig.hs} \Delta G_{hs} \left[1 - \frac{\left|\overrightarrow{r_{lig}} - \overrightarrow{r_{hs}}\right|}{R_{CO}}\right] \Theta\left(R_{CO} - \left|\overrightarrow{r_{lig}} - \overrightarrow{r_{hs}}\right|\right)$$

Equation 2. Displaced water - binding affinity relationship function. R_{CO} is the distance cutoff (2.24Å[8]), ΔG_{hs} is the free energy to transfer water from hydration site into bulk, and Θ is the Heaviside step function.

In a later study, Beuming et al. extended the implementation of WaterMap into the investigation of protein-peptide binding between the G-protein coupled receptors (GPCRs) and PDZ domains[71]. The PDZ domains exhibit a dominant hydrophobic interaction and only limited induce-fit structural changes, therefore PDZ domain binding should be an ideal model to validate the WaterMap protocol against the protein-peptide interaction. The results revealed that in Erbin[72] PDZ domain (PDB #1MFG), the

hydration site identified by the β-sheet and several polar amino acid side chains, such as Ser26, Arg49 and Gln51 was unfavorable because of the absence of potential hydrogen bond partners instead of being a hydrophilic site as people usually thought. Using WaterMap and the displaced solvent functional (Equation 2), the predicted free energy differences and the trends are in good agreement with their corresponding experimental data. However, like in protein-ligand models, the missing of entropy terms and strain energy made the predicted binding free energies more favorable than experimental measurements. Additionally, the missing of second and highest order terms in the calculation of entropy results in the protocol (Equation 1) overestimating $\Delta\Delta G$ between peptide pairs. Nonetheless, this protocol was indeed designed to predict binding free energy differences among congeneric ligands, instead of computing absolute binding free energies; therefore, the implementation of WaterMap protocol in specific type of protein-peptide binding models is still promising.

3. Recent Implementation of WaterMap

Displacing weak bound water molecules inside a protein binding pocket is usually a tricky, non-trivial business. Using computational approaches in lead optimization can foresee the difference in the water network between lead-enzyme complexes and apo structures, hence saving time, cost and effort. Accompany with the growing computer power, using the alchemical or PMF based sampling technique to predict the standard binding free energies has been reported. However, these Class 1 methods, as classified by Guvench and MacKerell[36], are still hardly feasible to be applied for a large amount of lead compounds. Under such a scenario, WaterMap and other Class 2 methods obviously have a huge advantage, especially for congeneric ligands sets where WaterMap works fantastically well. WaterMap has been applied to multiple pharmaceutically important enzyme targets, including kinase, GPCR and many others. In the following section, we will go over recent literature reports that have implemented this *in silico* protocol.

Robinson, Sherman and Farid[73] implemented WaterMap protocol to illustrate the thermodynamic profiles of water molecules in the binding sites of four kinase[74] systems. The results allowed them to rationalize the puzzled binding selectivity and improve structure-activity relationships (SARs). Pearlstein et al.[75] applied WaterMap to detect the

locations and investigate the thermodynamic properties of hydration sites in the binding of LDL receptor (LDL-R) into proprotein convertase subtilisin-kexin type 9 (PCSK9)[76-78]. Their computed energies of placing water in stable hydration site and water displacement from unstable hydration site showed good agreement with measured k_{ON}[79] and k_{OFF}[79], respectively. Their assumption that (in this case) the rate-determining step of protein-ligand binding and unbinding are associated with the unfavorable water displacement and resolvation process was validated. Higgs, Beuming and Sherman[80] mapped the hydration sites and studied the thermodynamic profiles for the GPCR adenosine A_{2A} (A2A) receptor[81-83] in order to explain the interesting and inexplicable SARs between triazolylpurine analogs and A2A receptor[6]. Based on their outputs, unfavorable small ligands displace the stable water, while favorable 'longer' candidates displace more thermodynamically unstable water as they extend themselves into that region, thus explaining the odd SARs previously observed. Once again, finding 'thermodynamically' correct displaceable water becomes the key to improve the potency in drug discovery. Abel, Salam and coworkers[84] extend the approach to explain the SARs established using traditional computational methods in the inhibitor binding in blood coagulation factor serine protease[85-87]. Small modifications to the ligands (e.g. chloro- to methyl-) lead to favorable water displacement in certain subpockets such as the ester-binding pocket, thus increasing the potency. The authors also reported their approach to apply WaterMap produced displacement energy into prediction of relative binding free energies, with additional terms included.

Although it is commonly accepted that hydrophobic effect accounts for a significant part of ligand-receptor binding affinities, it is yet clear whether there are multiple sources of hydrophobic effect. Recently, Snyder et al.[60] reported their investigation of the molecular recognition mechanism of arylsulfonamides binding to carbonic anhydrase. This work aimed to distinguish different types of hydrophobic effects in biological recognition process. Though most binding processes featured hydrophobic association are considered as entropically driven[88], the conversions of a couple of monocyclic ligands to corresponding bicyclic ligands exhibit an enthalpy-driven hydrophobic effect[88,89] ($\Delta\Delta G$ = -2.8 kcal/mol, $\Delta\Delta H$ = -3.0 kcal/mol, $-T\Delta\Delta S$ = 0.2 kcal/mol). The enthalpically unstable water molecules were displaced with the increase in

ligand size (volume), resulting in gain in affinity. The reordering of water network in the binding pocket was responsible for the enthalpy-entropy compensation while the increasing number of water caused the loss in entropy. Different thermodynamic profiles of hydrophobic effects were discussed in this study. It was also suggested that the 3-D mapping of water in binding pocket might be as important as the shape of the site itself.

Displacing thermodynamically unstable water against a β-sheet can govern the binding potency in PDZ domains[71]. Beuming et al.[90] went further to scan the hydration sites close to the surfaces of various proteins to generate the thermodynamic profiles and a general picture of these structured water molecules. It was found that waters around α-helices or β-sheets were more desirable to be displaced than those residing around loops. Although water molecules resolved in crystal structure always attract more attention, this work illustrated that the thermodynamic stability has little correlation with the degree of a water molecule buried in an enzyme. The physical underlining of hydrophobic effects was once again proved to be complex, but the hot spots detected by WaterMap usually imply thermodynamic instability and pharmaceutical druggability, while those cool spots usually require careful attention.

WaterMap does not need previous knowledge of the ligand in question, hence exhibits promise for effectively filtering new lead optimizations (in series of congeneric ligands). With the aid from WaterMap, Chrencik et al.[91] inspected the critical role of the nitrile group from CP-690550 in Janus kinases (JAKs) inhibiting[92,93]; Laha et al.[94] were able to develop and rank a new set of 2,4-diaminothiazoles against Cyclin-dependent kinase 5 (Cdk5)[95,96], and their new lead improved in vitro mouse microsomal stability; while Knegtel and Robinson[97] improved their virtual screening results in the examination of a set of interleukin-2 inducible T cell kinase (Itk) inhibitors[98] and identified a unique displacement of thermodynamically unfavorable water molecule with a favorably solvated aromatic ring nitrogen. Explicit water energy is usually ignored in most scoring functions, but the importance of this term to accurately predict ligand potency should not be not neglected. In the CADD against plasmodium by targeting falcipain (FP) cysteine protease[99,100], Shah et al.[101] found that water displacement energies could be used to better interpret puzzling SAR observations seen for FP-2 and FP-3 inhibitors search.

Implementing WaterMap analysis into traditional CADD process has successfully helped scientists to improve ranking leap optimization and establish better interpretation of SARs. However, on the other hand, efforts were also made to compare this protocol to other widely utilized methods, for instance MM-GB/SA, to try to address the advantage and limitations of WaterMap[102,103].

Guimaraes and Mathiowetz[103] reported the first attempt in this regard, in which they examined series of CKD2 and fXa inhibitors. The authors reasoned WaterMap predicted free energy associated with displacing water upon ligand binding and, therefore, could be used as a correction term to replace the protein desolvation term in GB to improve MM-GB/SA rescoring. Using MM-GB/SA with the GB desolvation term excluded, the authors were able to achieve R^2 of 0.75 and 0.65 for congeneric sets (thus they focused on relative binding free energy instead of absolute binding free energy) of fXa and CKD2 inhibitors, respectively. They improved the results to 0.71 and 0.68 accordingly using WaterMap. While the best results were obtained when integrating WaterMap predicted water displacement energy into the MM-GB/SA procedure (without the GB desolvation term. Although the margin is not quite large, it validated the idea that the two methods can be complementarily utilized in drug discovery. However, as the authors also noted, the lack of dynamic screening effect that causes overweighed electrostatic contribution in MM-GB/SA binding energies, could lead to unbalanced preference of hydrophobic ligand over small or polar molecules. Later, Kohlmann, Zhu and Dalgano[102] from ARIAD Pharmaceuticals performed a potency prediction study on an extensive congeneric series of small-molecule SRC tyrosine kinase inhibitors[104,105]. Although the enzyme flexibility was minimized in this study, WaterMap still failed to generate a rational correlation with experimental results when flexibility plays a role that cannot be ignored, this is due to the lack of terms such as protein-ligand interactions and/or ligand entropy. This case is more illustrative in terms of addressing the limitation of the WaterMap strategy: it is an efficient method for binding scenarios dominant by hydrophobic effects but an inappropriate quantitative scoring tool when hydrophobic effects were not the main driving force.

4. Expert Opinion

In this section, we will first discuss the potential improvements that can be made to the WaterMap protocol, and our advices on how to more effectively implement this popular method and, and follow with extending discussions to treating water molecules in leap optimization stage.

The WaterMap procedure is built on clustering based analysis followed by a short restraint MD simulation[7,8]. The OPLS-2005 all-atom force field[64,65] and TIP4P[18] 4-site water model were used for MD set up. The goal of this protocol is to isolate hydration sites and scan their thermodynamic profiles; therefore, the choice of water model is of great importance and our discussion will start here. Comparing to more commonly used 3-site TIP3P[18] model, TIP4P presents improvements on several aspects, such as diffusion coefficient and density, at a modest increase in computational cost[106,107]. However, TIP4P model generally produces poor Lennard-Jones potential well (ε), whereas TIP5P[17] model developed by Mahoney and Jorgensen well reproduces this property. More importantly, TIP5P is the first model that exhibits maximum density at 4 degree (277 K) and it also reasonably reproduced pressure effects. Considering these factors, it seems taking the advantage of TIP5P over TIP4P could benefit the WaterMap analysis. Unfortunately, TIP5P is much more computer time demanding than TIP4P, not to mention TIP4P outperforms TIP5P on reproducing liquid-gas (vapor) equilibration[24]. Modified TIP4P models have been introduced, by either fitting to reproduce maximum density at 4 degree (TIP4P/2005[108]), or introducing polarization effect (TIP4P-FQ[109]), or refitting to reproduce other properties[110,111]. The implementation of these water models could possibly improve the WaterMap results, though the importance of these add-on properties to the simulation of solvation effect is yet clear. But, as faced in other computer approaches, this is again a problem of balance between computational expense and accuracy.

WaterMap has been integrated into Glide[61,112], available in the Schrodinger suite of programs[113] and the implementation is not difficult (may be subjected to service fees). For users without access to Schrodinger suite of programs, the analysis strategy for MD simulations introduced by Young et al.[7] and Abel et al.[8] is available freely, though it is much less straightforward. However, as mentioned in the original proposal[7] and discussed in a recent validation study[102], the hydrophobic enclosure is of the greatest

importance to the success implementation of WaterMap analysis. In other words, the output of this analysis is highly sensitive to the topology of the system. Considering this fact, using this procedure against structures known to exhibit conformational flexibilities would be inappropriate. An educated guess or judgment should be made before implementing WaterMap to make sure that the principle driving force of the binding is hydrophobic effect thus fitting the criteria of this protocol. In addition, the advertisement of WaterMap protocol emphasis the use of a single structure (for the sake of computational efficiency). That would lead to the outputs highly sensitive to the quality of the choice of starting structure, whereas running multiple starting structures would significantly increase the computing cost. Moreover, it is critical to keep in mind that WaterMap might fail to capture the chemistry of ligands other than those congeneric ligands due to the lack of term describing ligand conformational entropy, ligand-protein interactions and many others. To summarize, researchers should conduct some prep work to determine whether the ligand, structural and topological factors are all favored the choice of this protocol.

Although WaterMap is able to give useful insights in the ligand-receptor chemistry that other methods failed to, it is far from a robust approach that can handle the binding process by itself. As shown in attempts made by different groups[102,103], the combination of WaterMap and MM-GB/SA displayed improving results to both end. However, this approach, though promising, is not guaranteed as well, as suggested by Kohlmann et al.[102]. Investigators also need a pre-judgment to ensure the topological factor of the system suitable is for the WaterMap approach. Possible strategies suitable for this pre-judgment including calculating non-polar surface area, estimating the fraction of polar surface area, or simply visually checking the binding site surrounding environment if experienced. Comparing to MM-GB/SA (or MM-PB/SA), FEP (and other computer time demanding free energy approaches) is/are more robust in determining the ligand absolute binding free energies, but are also structurally dependent[56,57]. Though FEP does allow structural flexibility, Michel et al. showed the results were sensitive to the initial water placement and the Jorgensen lab has applied their JAWS[56] to improve the water placement in the initial structures that feed FEP simulations[25]. An interesting (and of course, computer time demanding) plot is to use the hydration sites determined through

WaterMap approach as the water placement feeding to subsequent FEP simulations. Useful insights could be expected and the development of improved scoring function might be benefit from this attempt. However, as mentioned repeatedly, this is again a balance between projected computer expense and anticipated accuracy, as faced everywhere by computational chemists.

Lead optimization is an important stage in drug discovery responsible for fulfilling the requirement for clinical usefulness and water displacement represents a big challenge in this process. More careful, physics-based approaches upon the computational power permits should be made in order for successes. Combining single-point energy methods such as MM-GB/SA rescoring with WaterMap explicit water analysis made a big strides along this path, more robust (and also efficient) methods/approaches are still in need to fully govern the protein-ligand binding predictions.

Declaration of Interest

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Also, we should acknowledge support from the LDRD SI: 12-SI-004.

(1) Warshel, A. *P Natl Acad Sci USA* **1978**, *75*, 5250.

(2) Dill, K. A. *Biochemistry-Us* **1990**, *29*, 7133.

(3) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.

(4) Wang, Y.; Schlick, T. *J Am Chem Soc* **2008**, *130*, 13240.

(5) Yang, Y.; Miao, Y. P.; Wang, B.; Cui, G. L.; Merz, K. M. *Biochemistry-Us* **2012**, *51*, 2606.

(6) Jacobson, K. A.; Gao, Z. G. *Nat Rev Drug Discov* **2006**, *5*, 247.

(7) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. *P Natl Acad Sci USA* **2007**, *104*, 808.

(8) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J Am Chem Soc* **2008**, *130*, 2817.

(9) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv Drug Deliver Rev* **2001**, *46*, 3.

(10) Leach, A. R. *Molecular Modelling: Principles and Applications*; 2nd ed.; Prentice Hall: New York, 2001.

(11) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *J Phys Chem B* **1997**, *101*, 1190.

(12) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J Phys Chem B* **1998**, *102*, 10983.

(13) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Accounts Chem Res* **2000**, *33*, 889.

(14) Kuhn, B.; Kollman, P. A. *J Am Chem Soc* **2000**, *122*, 3909.

(15) Bashford, D.; Case, D. A. *Annu Rev Phys Chem* **2000**, *51*, 129.

(16) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275.

(17) Mahoney, M. W.; Jorgensen, W. L. *J Chem Phys* **2000**, *112*, 8910.

(18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* **1983**, *79*, 926.

(19) Mark, P.; Nilsson, L. *J Phys Chem A* **2001**, *105*, 9954.

(20) Amira, S.; Spangberg, D.; Hermansson, K. *Chem Phys* **2004**, *303*, 327.

(21) Lobaugh, J.; Voth, G. A. *J Chem Phys* **1997**, *106*, 2400.

(22) Borgis, D.; Staib, A. *Chem Phys Lett* **1995**, *238*, 187.

(23) Paesani, F.; Voth, G. A. *J Phys Chem B* **2009**, *113*, 5702.

(24) Lisal, M.; Kolafa, J.; Nezbeda, I. *J Chem Phys* **2002**, *117*, 8892.

(25) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J Am Chem Soc* **2009**, *131*, 15403.

(26) Rezai, T.; Bock, J. E.; Zhou, M. V.; Kalyanaraman, C.; Lokey, R. S.; Jacobson, M. P. *J Am Chem Soc* **2006**, *128*, 14073.

(27) Berne, B. J.; Weeks, J. D.; Zhou, R. H. *Annu Rev Phys Chem* **2009**, *60*, 85.

(28) Ladbury, J. E. *Chem Biol* **1996**, *3*, 973.

(29) Hummer, G. *Nat Chem* **2010**, *2*, 906.

(30) Cheung, M. S.; Garcia, A. E.; Onuchic, J. N. *P Natl Acad Sci USA* **2002**, *99*, 685.

(31) Lazaridis, T.; Karplus, M. *Biophys Chem* **2003**, *100*, 367.

(32) Dougan, L.; Feng, G.; Lu, H.; Fernandez, J. M. *P Natl Acad Sci USA* **2008**, *105*, 3185.

(33) Smith, L.; Cheetham, A. K.; Morris, R. E.; Marchese, L.; Thomas, J. M.; Wright, P. A.; Chen, J. *Science* **1996**, *271*, 799.

(34) Xu, D. G.; Guo, H. *Febs Lett* **2005**, *579*, 4249.

(35) Deng, Y. Q.; Roux, B. *J Phys Chem B* **2009**, *113*, 2234.

(36) Guvench, O.; MacKerell, A. D. *Curr Opin Struc Biol* **2009**, *19*, 56.

(37) Mobley, D. L.; Dill, K. A. *Structure* **2009**, *17*, 489.

(38) Wereszczynski, J.; McCammon, J. A. *Q Rev Biophys* **2012**, *45*, 1.

(39) Lu, Y. P.; Wang, R. X.; Yang, C. Y.; Wang, S. M. *J Chem Inf Model* **2007**, *47*, 668.

(40) Wong, S. E.; Lightstone, F. C. *Expert Opin Drug Dis* **2011**, *6*, 65.

(41) Huang, N.; Shoichet, B. K. *J Med Chem* **2008**, *51*, 4862.

(42) Robeits, B. C.; Mancera, R. L. *J Chem Inf Model* **2008**, *48*, 397.

(43) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. E. *J Med Chem* **2005**, *48*, 2308.

(44) Thilagavathi, R.; Mancera, R. L. *J Chem Inf Model* **2010**, *50*, 415.

(45) Santos, R.; Hritz, J.; Oostenbrink, C. *J Chem Inf Model* **2010**, *50*, 146.

(46) Bottoms, C. A.; White, T. A.; Tanner, J. J. *Proteins* **2006**, *64*, 404.

(47) Goodford, P. J. *J Med Chem* **1985**, *28*, 849.

(48) Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. *J Mol Biol* **2006**, *358*, 289.

(49) Garcia-Sosa, A. T.; Mancera, R. L.; Dean, P. M. *J Mol Model* **2003**, *9*, 172.

(50) Pan, C. F.; Mezei, M.; Mujtaba, S.; Muller, M.; Zeng, L.; Li, J. M.; Wang, Z. Y.; Zhou, M. M. *J Med Chem* **2007**, *50*, 2285.

(51) Mccammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.

(52) Yu, H. T.; Rick, S. W. *J Am Chem Soc* **2009**, *131*, 6608.

(53) Yu, H. T.; Rick, S. W. *J Phys Chem B* **2010**, *114*, 11552.

(54) Baron, R.; Setny, P.; McCammon, J. A. *J Am Chem Soc* **2010**, *132*, 12091.

(55) Baron, R.; Molinero, V. *Journal of Chemical Theory and Computation* **2012**.

(56) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J Phys Chem B* **2009**, *113*, 13337.

(57) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2010**, *6*, 3850.

(58) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. *J Am Chem Soc* **2007**, *129*, 2577.

(59) Michel, J.; Essex, J. W. *J Comput Aid Mol Des* **2010**, *24*, 639.

(60) Snyder, P. W.; Mecinovic, J.; Moustakas, D. T.; Thomas, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; Heroux, A.; Sherman, W.; Whitesides, G. M. *P Natl Acad Sci USA* **2011**, *108*, 17889.

(61) ; Schro□dinger, Inc.

(62) Lazaridis, T. *J Phys Chem B* **1998**, *102*, 3542.

(63) Lazaridis, T. *J Phys Chem B* **1998**, *102*, 3531.

(64) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J Phys Chem B* **2001**, *105*, 6474.

(65) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J Am Chem Soc* **1996**, *118*, 11225.

(66) Padmanabhan, K.; Padmanabhan, K. P.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D. T.; Cardin, A. D.; Kisiel, W. *J Mol Biol* **1993**, *232*, 947.

(67) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. *Biophys J* **2000**, *79*, 2966.

(68) Carugo, O.; Bordo, D. *Acta Crystallogr D* **1999**, *55*, 479.

(69) Nakasako, M.; Takahashi, H.; Shimba, N.; Shimada, I.; Arata, Y. *J Mol Biol* **1999**, *291*, 117.

(70) Mattos, C. *Trends Biochem Sci* **2002**, *27*, 203.

(71) Beuming, T.; Farid, R.; Sherman, W. *Protein Sci* **2009**, *18*, 1609.

(72) Skelton, N. J.; Koehler, M. F. T.; Zobel, K.; Wong, W. L.; Yeh, S.; Pisabarro, M. T.; Yin, J. P.; Lasky, L. A.; Sidhu, S. S. *J Biol Chem* **2003**, *278*, 7645.

(73) Robinson, D. D.; Sherman, W.; Farid, R. *Chemmedchem* **2010**, *5*, 618.

(74) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912.

(75) Pearlstein, R. A.; Hu, Q. Y.; Zhou, J.; Yowe, D.; Levell, J.; Dale, B.; Kaushik, V.; Daniels, D.; Hanrahan, S.; Sherman, W.; Abel, R. *Abstr Pap Am Chem S* **2010**, *240*.

(76) Horton, J. D.; Cohen, J. C.; Hobbs, H. H. *Trends Biochem Sci* **2007**, *32*, 71.

(77) Zhang, D. W.; Lagace, T. A.; Garuti, R.; Zhao, Z.; McDonald, M.; Horton, J. D.; Cohen, J. C.; Hobbs, H. H. *J Biol Chem* **2007**, *282*, 18602.

(78) Horton, J. D.; Cohen, J. C.; Hobbs, H. H. *J Lipid Res* **2009**, *50*, S172.

(79) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. *Nat Rev Drug Discov* **2006**, *5*, 730.

(80) Higgs, C.; Beuming, T.; Sherman, W. *Acs Med Chem Lett* **2010**, *1*, 160.

(81) Libert, F.; Parmentier, M.; Lefort, A.; Dinsart, C.; Vansande, J.; Maenhaut, C.; Simons, M. J.; Dumont, J. E.; Vassart, G. *Science* **1989**, *244*, 569.

(82) Libert, F.; Passage, E.; Parmentier, M.; Simons, M. J.; Vassart, G.; Mattei, M. G. *Genomics* **1991**, *11*, 225.

(83) Libert, F.; Passage, E.; Parmentier, M.; Simons, M. J.; Vassart, G.; Mattei, M. G. *Genomics* **1994**, *23*, 305.

(84) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. *Chemmedchem* **2011**, *6*, 1049.

(85) Hedstrom, L. *Chem Rev* **2002**, *102*, 4501.

(86) Leung, D.; Abbenante, G.; Fairlie, D. P. *J Med Chem* **2000**, *43*, 305.

(87) Katz, B. A.; Sprengeler, P. A.; Luong, C.; Verner, E.; Elrod, K.; Kirtley, M.; Janc, J.; Spencer, J. R.; Breitenbucher, J. G.; Hui, H.; McGee, D.; Allen, D.; Martelli, A.; Mackman, R. L. *Chem Biol* **2001**, *8*, 1107.

(88) Setny, P.; Baron, R.; McCammon, J. A. *Journal of Chemical Theory and Computation* **2010**, *6*, 2866.

(89) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew Chem Int Edit* **2003**, *42*, 1210.

(90) Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. *Proteins* **2012**, *80*, 871.

(91) Chrencik, J. E.; Patny, A.; Leung, I. K.; Korniski, B.; Emmons, T. L.; Hall, T.; Weinberg, R. A.; Gormley, J. A.; Williams, J. M.; Day, J. E.; Hirsch, J. L.; Kiefer, J. R.; Leone, J. W.; Fischer, H. D.; Sommers, C. D.; Huang, H. C.; Jacobsen, E. J.; Tenbrink, R. E.; Tomasselli, A. G.; Benson, T. E. *J Mol Biol* **2010**, *400*, 413.

(92) Firmbachkraft, I.; Byers, M.; Shows, T.; Dallafavera, R.; Krolewski, J. J. *Oncogene* **1990**, *5*, 1329.

(93) Partanen, J.; Makela, T. P.; Alitalo, R.; Lehvaslaiho, H.; Alitalo, K. *P Natl Acad Sci USA* **1990**, *87*, 8913.

(94) Laha, J. K.; Zhang, X. M.; Qiao, L. X.; Liu, M.; Chatterjee, S.; Robinson, S.; Kosik, K. S.; Cuny, G. D. *Bioorg Med Chem Lett* **2011**, *21*, 2098.

(95) Malumbres, M.; Barbacid, M. *Trends Biochem Sci* **2005**, *30*, 630.

(96) Peterson, D. W.; Ando, D. M.; Taketa, D. A.; Zhou, H. J.; Dahlquist, F. W.; Lew, J. *P Natl Acad Sci USA* **2010**, *107*, 2884.

(97) Knegtel, R. M. A.; Robinson, D. D. *Mol Inform* **2011**, *30*, 950.

(98) Charrier, J. D.; Miller, A.; Kay, D. P.; Brenchley, G.; Twin, H. C.; Collier, P. N.; Ramaya, S.; Keily, S. B.; Durrant, S. J.; Knegtel, R. M. A.; Tanner, A. J.; Brown, K.; Curnock, A. P.; Jimenez, J. M. *J Med Chem* **2011**, *54*, 2341.

(99) Dluzewski, A. R.; Rangachari, K.; Wilson, R. J. M.; Gratzer, W. B. *Exp Parasitol* **1986**, *62*, 416.

(100) Fidock, D. A. *Nature* **2010**, *465*, 297.

(101) Shah, F.; Gut, J.; Legac, J.; Shivakumar, D.; Sherman, W.; Rosenthal, P. J.; Avery, M. A. *J Chem Inf Model* **2012**, *52*, 696.

(102) Kohlmann, A.; Zhu, X. T.; Dalgarno, D. *Acs Med Chem Lett* **2012**, *3*, 94.

(103) Guimaraes, C. R. W.; Mathiowetz, A. M. *J Chem Inf Model* **2010**, *50*, 547.

(104) Dalgarno, D.; Stehle, T.; Narula, S.; Schelling, P.; van Schravendijk, M. R.; Adams, S.; Andrade, L.; Keats, J.; Ram, M.; Jin, L.; Grossman, T.; MacNeil, I.; Metcalf, C.; Shakespeare, W.; Wang,

Y.; Keenan, T.; Sundaramoorthi, R.; Bohacek, R.; Weigele, M.; Sawyer, T. *Chem Biol Drug Des* **2006**, *67*, 46.

(105) Shakespeare, W. C.; Wang, Y.; Bohacek, R.; Keenan, T.; Sundaramoorthil, R.; Metcalf, C.; Dilaurol, A.; Roeloffzen, S.; Liu, S. Y.; Saltmarsh, J.; Paramanathan, G.; Dalgarnol, D.; Narulal, S.; Pradeepan, S.; van Schravendijk, M. R.; Keats, J.; Ram, M.; Liou, S.; Adams, S.; Wardwell, S.; Bogus, J.; Luliucci, J.; Weigele, M.; Xing, L. P.; Boyce, B.; Sawyer, T. K. *Chem Biol Drug Des* **2008**, *71*, 97.

(106) Guillot, B. *J Mol Liq* **2002**, *101*, 219.

(107) Hess, B.; van der Vegt, N. F. A. *J Phys Chem B* **2006**, *110*, 17616.

(108) Abascal, J. L. F.; Vega, C. *J Chem Phys* **2005**, *123*.

(109) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J Chem Phys* **1994**, *101*, 6141.

(110) Lawrence, C. P.; Skinner, J. L. *Chem Phys Lett* **2003**, *372*, 842.

(111) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J Chem Phys* **2004**, *120*, 9665.

(112) ; Schro□dinger, Inc.

(113) Repasky, M. P.; Murphy, R. B.; Banks, J. L.; Greenwood, J. R.; Tubert-Brohman, I.; Bhat, S.; Friesner, R. A. *J Comput Aid Mol Des* **2012**, *26*, 787.